Effectiveness of physiotherapy for lateral epicondylitis: a systematic review

Smidt, N.; Assendelft, W.J.J.; Arola, H.; Malmivaara, A.; Green, S.; Buchbinder, R.; van der Windt, D.A.W.M.; Bouter, L.M.

published in
Annals of Medicine
2003

DOI (link to publisher)
10.1080/07853890310004138

document version
Publisher's PDF, also known as Version of record

◆ ORIGINAL ARTICLE ◆

Taylor & Francis healthsciences

Effectiveness of physiotherapy for lateral epicondylitis: a systematic review

Nynke Smidt[1], Willem JJ Assendelft[2], Heikki Arola[3], Antti Malmivaara[4], Sally Green[5], Rachelle Buchbinder[6], Daniëlle AWM van der Windt[1] and Lex M Bouter[1]

AIM: To evaluate the available evidence of the effectiveness of physiotherapy for lateral epicondylitis of the elbow.
METHOD: Randomised controlled trials (RCTs) identified by a highly sensitive search strategy in six databases in combination with reference checking. Two independent reviewers selected RCTs that included a physiotherapy intervention, patients with lateral epicondylitis, and at least one clinically relevant outcome measure. No language restrictions were made. Methodological quality was independently assessed by two blinded reviewers. A best evidence synthesis, including a quantitative and qualitative analysis, was conducted, weighting the studies with respect to their internal validity, statistical significance, clinical relevance, and statistical power.
RESULTS: 23 RCTs were included in the review, evaluating the effects of lasertherapy, ultrasound treatment, electrotherapy, and exercises and mobilisation techniques. Fourteen studies satisfied at least 50% of the internal validity criteria. Except for ultrasound, pooling of data from RCTs was not possible because of insufficient data, or clinical or statistical heterogeneity. The pooled estimate of the treatment effects of two studies on ultrasound compared to placebo ultrasound, showed statistically significant and clinically relevant differences in favour of ultrasound. There is insufficient evidence either to demonstrate benefit or lack of effect of lasertherapy, electrotherapy, exercises and mobilisation techniques for lateral epicondylitis.
CONCLUSIONS: Despite the large number of studies, there is still insufficient evidence for most physiotherapy interventions for lateral epicondylitis due to contradicting results, insufficient power, and the low number of studies per intervention. Only for ultrasound, weak evidence for efficacy was found. More better designed, conducted and reported RCTs are needed.

Keywords: meta-analysis; physical therapy; randomised controlled trials; tendinitis; tennis elbow

Ann Med 2003; 35: 51–62

Introduction

Lateral epicondylitis (tennis elbow) is a common medical problem. The annual incidence of this disorder is between 1% and 3% in the general population (1–3). The average duration of a typical episode of lateral epicondylitis is reported to be between 6 months and 2 years (4).

A wide array of physiotherapy interventions is used for treating lateral epicondylitis. Choices seem to be driven by tradition or are based on trials with a relatively small study size or poor methodological quality (5). In an attempt to systematically summarise the available evidence Labelle et al. (6) intended to perform a quantitative meta-analysis of various treatments for lateral epicondylitis, including five randomised controlled trials (RCTs) of physiotherapy. Because of poor methodological quality and the contradictory results of the studied trials, the authors concluded that there was insufficient scientific evidence for any particular type of treatment for lateral epicondylitis (6).

The review by Labelle et al. (6) only covered the RCTs indexed in MEDLINE and EMBASE during

[1]Institute for Research in Extramural Medicine, VU University Medical Center, Amsterdam, The Netherlands, [2]Dutch College of General Practitioners (NHG), Department of Guideline Development, Utrecht, The Netherlands, [3]Tampere Occupational Health Centre, Tampere, Finland, [4]Finnish Institute for Occupational Health, Helsinki, Finland, [5]Monash Institute of Public Health, Monash University, Melbourne, Australia, [6]Department of Epidemiology and Preventive Medicine, Monash University, Cabrini Medical Centre, Melbourne, Australia.

Correspondence to: Ms. N. Smidt PhD, Institute for Research in Extramural Medicine, VU University Medical Center, Van der Boechorststraat 7, NL-1081 BT, Amsterdam, The Netherlands. Fax: +31 20 444 8181; E-mail: n.smidt.emgo@med.vu.nl

 ISSN 0785-3890

1966–1990, and included only studies published in French or English. Other systematic reviews have only addressed the effectiveness of ultrasound for tennis elbow (7, 8). According to the current science of systematic reviews, a more comprehensive search strategy is clearly indicated (9–11). In addition, refraining from statistical pooling is only one of the options available for dealing with the insufficient methodological quality of RCTs (12).

Consequently, we decided to perform an updated, more comprehensive systematic review of the effectiveness of physiotherapy for lateral epicondylitis, using explicit and systematic methods for the search, selection and quality assessment of RCTs according to the state of the art and aiming at statistical pooling of subsets of comparable studies. The objective of our systematic review is to determine the effectiveness of physiotherapy in patients with lateral epicondylitis based on clinically relevant outcome measures, such as pain, global improvement and grip strength.

Key messages

- There is still insufficient evidence to draw firm conclusions on the effects of laser, electrotherapy, exercises and mobilisation techniques for lateral epicondylitis.
- Weak evidence was found for the effectiveness of ultrasound compared to placebo ultrasound for lateral epicondylitis.
- Lack of evidence for the effectiveness of physiotherapeutic interventions is mainly the result of low statistical power, inadequate internal validity and insufficient reported data of the majority of trials.

Materials and methods

Searching

One reviewer (NS) searched computerised bibliographical databases (MEDLINE 01/1966 — 01/1999, EMBASE 01/1988 — 01/1999, and CINAHL 01/1982 — 01/1999) without language restrictions (10, 13), using the highly sensitive Cochrane Collaboration search strategy, which aims to identify all randomised controlled trials (14, 15).

Additional specific subject headings and free text words were used to identify papers on lateral epicondylitis and physiotherapy. The Cochrane Controlled Trial Register was searched (14). An additional search for systematic reviews on our topic was carried out in EMBASE and MEDLINE (9). Furthermore, the Current Contents (July 1999) database was searched, and a computer-aided search was carried out in the trial register of the Cochrane field of 'Rehabilitation and related therapies'. Finally, references from retrieved articles were screened (citation tracking). To determine whether a study should be included, the abstracts of all identified hits were assessed by two reviewers (NS and WJJA) independently. If there was any doubt, the full article was retrieved and read independently by both reviewers. Disagreements were discussed in a consensus meeting.

Selection

For this systematic review we included studies that met the following conditions: 1) Patients with lateral epicondylitis, or lateral elbow pain increased by pressure on the lateral epicondyle and during resisted dorsiflexion of the wrist; 2) At least one of the treatments included a physiotherapy intervention. Physiotherapy had to be contrasted with either placebo, no treatment or another conservative treatment. Studies comparing physiotherapy with surgical treatment were excluded; 2) Treatment regimens were allocated by a random procedure. The words 'random' or 'randomised' had to be mentioned in the article (16); 3) At least one clinically relevant outcome measure (pain, global improvement, elbow specific functional status, grip strength, or sick leave) was included; 5) Follow-up was at least 1 day; 6) Published as a full report before July 1999.

Quality assessment

The Amsterdam-Maastricht consensus list (8, 17) was used for methodological quality assessment, consisting of internal validity criteria, descriptive criteria and statistical criteria (Table 1). The descriptive criteria refer to the external validity (generalisability) of the study and are used to identify homogeneous subgroups and conduct sensitivity analyses. Our criteria list includes all criteria of the widely used lists of Jadad et al. (18) and Verhagen et al. (19).

A research assistant not involved in any other component of the systematic review (SK, see acknowledgements), blinded all articles eligible for the review for authors, journal and year of the trial. Included articles were independently assessed for methodological quality by two blinded reviewers (NS and HA). Initial disagreement was evaluated and expressed as percentage of agreement and kappa statistics (20, 21). In a consensus meeting disagreements were discussed and resolved. For studies published in languages other than English, German or Dutch, the help of a native speaker or translator with content expertise was obtained (see acknowledgements). Each methodological quality criterion

 ISSN 0785-3890

Table 1. Criteria for the methodological assessment of randomised clinical trials *

Internal validity criteria	
V1	Adequate randomisation: adequate procedure for generation of a random number sequence
V2	Concealed randomisation
V3	Baseline similarity of intervention group
V4	Control for co-interventions in design
V5	Co-interventions reported for each group separately
V6	Adherence to interventions: >70% for each group (index and reference group)
V7	Care provider blinded
V8	Patient blinded
V9	Withdrawals and drop-outs: ≤20% for short term follow-up, and ≤30% for intermediate term and long term follow-up and no substantial bias (numerical inequality between groups or differences in reasons for withdrawal/drop-out)
V10	Identical timing of outcome assessment
V11	Intention-to-treat analysis
V12	Outcome assessor blinded
Descriptive criteria	
D1	Specification of eligibility criteria
D2	Baseline characteristics described
D3	Description of interventions
D4	Adverse effects described and attributed to allocated treatment, or explicit report of 'no adverse effects'
D5	Short-term follow-up (≤6 weeks)
D6	Intermediate-term follow-up (6 weeks to 6 months)
D7	Long-term follow-up (≥6 months)
Statistical criteria	
S1	Presentation of sample size at randomisation and at follow-up
S2	Presentation of point estimates and distribution measures

* Operationalisation of the criteria are available on request from first author.

was rated positive, negative or inconclusive (insufficient information presented). A total score for the internal validity of each study was calculated, by summing up the number of positive criteria (range 0–12), higher scores indicating a lower likelihood of bias. The two blinded reviewers (NS and HA) independently extracted the data regarding the interventions, timing of outcome assessment, loss to follow-up and results.

Quantitative data synthesis

The results of each RCT were expressed as relative risks (RR) with corresponding 95% confidence interval for dichotomous data (14), a relative risk smaller than 1.0 indicating a beneficial effect of physiotherapy (14). For continuous data the standardised mean difference (SMD) was calculated: $SMD = (0_r - 0_t)/PSD$, where 0_r = mean improvement in the reference group, 0_t = mean improvement in the treatment group, and PSD = pooled standard deviation (14), SMDs less than zero indicating a beneficial effect in favour of physiotherapy. For studies that did not report the standard deviation of the difference in mean improvement, the SMD was calculated by conversion of the reported *P* value to a *t*-statistic with appropriate degrees of freedom (22). Confidence intervals of 95% were computed for the SMD. The SMD was interpreted as described by Cohen (23); i.e., an SMD of 0.2 was considered a small beneficial effect, 0.5 a medium effect, and 0.8 a large effect. Power estimates for an SMD of 0.5 and 0.8 were made, to explore whether trials could detect an existing difference between interventions with a significance level of 0.05. According to the power calculation ($\beta = 0.80$, $\alpha = 0.05$), at least 64 patients are needed in each group to detect a clinically relevant difference (SMD of 0.5) between the interventions. Relative risks were considered clinically relevant, if RR is smaller than 0.7 or larger than 1.5, thus in favour of the index or reference group, respectively.

Pre-planned stratified analyses were aimed at: short-term (≤6 weeks after randomisation), intermediate (6 weeks to 6 months), and long-term follow-up (≥6 months); different control groups (placebo, no treatment, physiotherapeutic treatment and other conservative treatment); acceptable and low internal validity scores (24) (cut-off point for acceptable internal validity was 7 or more (≥50% of total score)), and different physiotherapy interventions. In addition, we attempted to assess the influence of the following potential prognostic factors: presence of concomitant neck or shoulder pain and duration of elbow complaints at baseline.

Best evidence synthesis

Studies were weighted as to their internal validity, statistical significance, clinical relevance, and power. Decision rules to distinguish between 'strong', 'weak' and 'insufficient' evidence for the effectiveness of a physiotherapy intervention or for no differences in

 ISSN 0785-3890

effect are presented in Figure 1. Statistical pooling (quantitative analysis) using random effects model, was conducted on the following conditions. Firstly, studies had acceptable internal validity scores (24) (cut-off point for acceptable internal validity was 7 or more (>50% of the total score)), secondly, studies were clinically homogenous, and finally, statistical homogeneity (Chi square test $P > 0.05$) between studies existed. Clinically homogeneity between studies existed, if studies were comparable to the timing of outcome assessment ((short-term ($\leq$6 weeks), intermediate-term (6 weeks to 6 months), and long-term ($\geq$6 months) follow-up)), control group (no treatment, placebo, other physiotherapeutic treatment, and other conservative treatment), and outcome measure (e.g. pain, global improvement).

If a quantitative analysis of data was not possible, conclusions regarding the strength of evidence were based on the consistency of findings between individual studies (qualitative analysis). Results were considered consistent if more than 75% of the studies reported similar results on the same outcome measure (i.e. favouring the same intervention) (8, 15). In case statistical pooling was only possible for less than 75% of the studies, an additional qualitative analysis was performed. If different conclusions were found between quantitative and qualitative analysis, conclusions were based on the analysis with the strongest evidence.

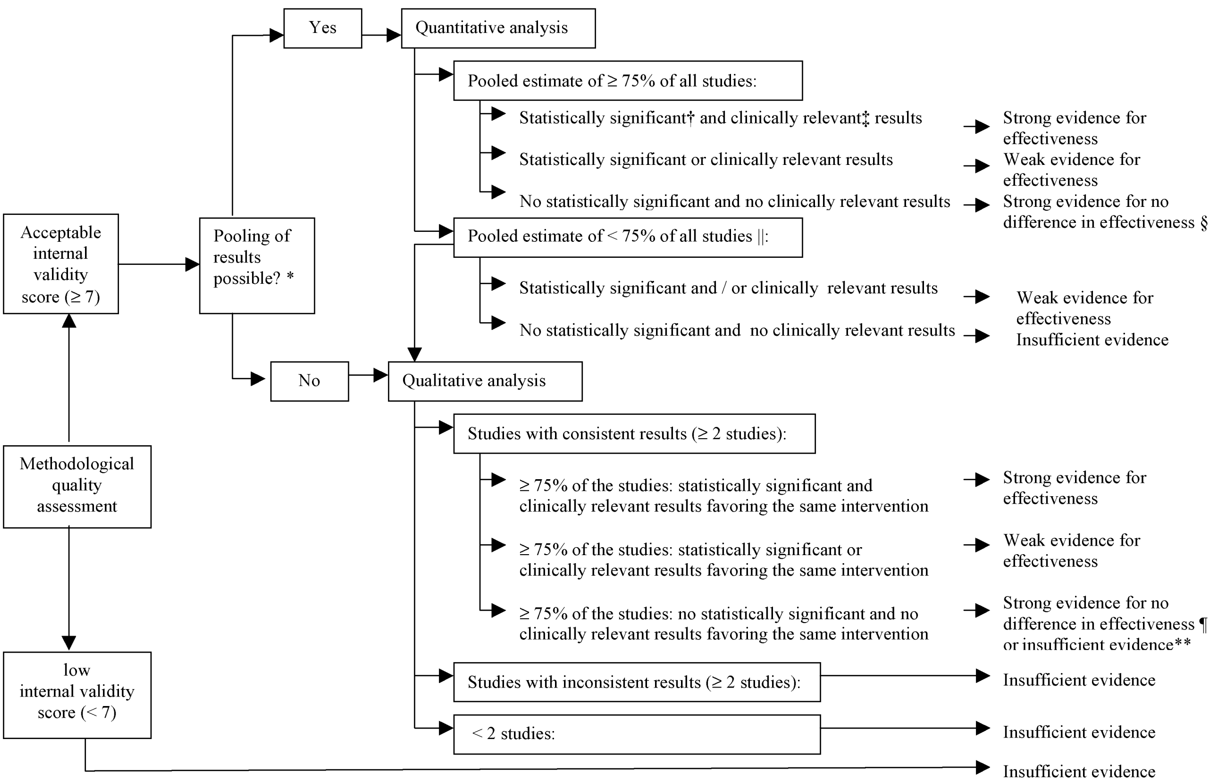
Results

Study selection

The results of our search strategy are presented in Figure 2. Out of a total of 1001 potentially relevant abstracts, 55 English and 15 non-English articles were considered to be potentially eligible for the review. Reviewing the full text resulted in inclusion of 23 articles (25–47). No additional eligible RCTs were found by screening the references. When indicated, the help of a native speaker was sought.

Study characteristics

Details regarding the study characteristics of included studies are available on request from the first author. Nine studies examined the effectiveness of ultrasound therapy (25, 28, 33, 35, 36, 40, 42–44), nine studies examined the effectiveness of laser therapy (30–32, 34, 38, 39, 41, 44, 45), while five studies focussed on



* Pooling of results: Statistical and clinical homogeneity between studies; † Statistically significant: *P*-value < 0.05; ‡ Clinically relevant: for dichotomous outcomes: RR ≤ 0.7 in favour of index group or ≥ 1.5 in favour of reference group, and for continuous outcomes: SMD ≤ -0.5 in favour of indexgroup or ≥ 0.5 in favour of reference group; § in case of sufficient power to detect an effect size of 0.5; || additional qualitative analysis will be performed. If different conclusions were found between quantitative and qualitative analysis, conclusions were based on the analysis with the strongest evidence; ¶ in case of sufficient power to detect an effect size of 0.5 in all studies; ** in case of insufficient power to detect an effect size of 0.5 in all studies.

Figure 1. Best-evidence synthesis.

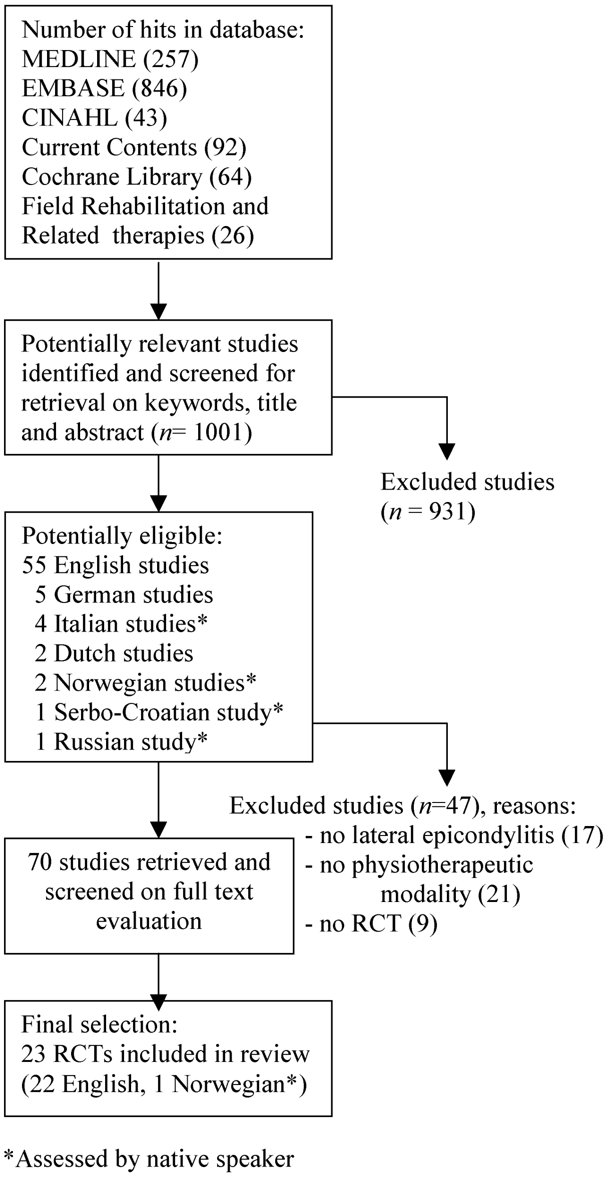


Figure 2. Study selection (48).

exercises or mobilisation techniques (28, 29, 42, 46, 47), and four studies on electrotherapy (26, 27, 35, 37).

Eleven studies (28, 32, 33, 34, 37, 38, 40, 42, 43, 45, 47) explicitly excluded patients with concomitant neck or shoulder complaints, four studies (36, 39, 41, 44) excluded only patients with concomitant neck complaints, seven studies (25–31, 35) did not specify whether patients had concomitant neck or shoulder complaints, and one (46, 49) included patients with concomitant neck and shoulder complaints.

All studies, with the exception of Pienimäki (42), performed a short-term outcome assessment ($\leq$6 weeks). Only seven studies (25, 31–34, 36, 46) included a long-term follow-up ($\geq$6 months).

Most studies included a mixed population of patients with acute, subacute and chronic lateral epicondylitis (25, 29–34, 36, 38, 40, 41, 43–45, 47). Five studies (26, 27, 37, 39, 42) restricted their inclusion criteria to patients with chronic lateral epicondylitis ($\geq$13 weeks), while three studies (28, 35, 46) did not specify the study population.

Assessment of outcome

Pain and grip strength were reported in most trials (15 and 19, respectively), although most studies did not perform a blinded outcome assessment (Table 2). Pain was measured either by visual analogue scales (25, 36, 39–45), ordinal scales varying between 3 and 9 points (29, 38, 46), or the outcome measure was not specified (32, 35, 47). Fourteen studies reported the global improvement or patient satisfaction, but a great variety of outcome measures were used.

Quality assessment

Table 3 presents the internal validity score for each study and the criteria for which bias was considered

Table 2. Outcome measures and blinded outcome assessment in 23 RCTs of physiotherapy for lateral epicondylitis.

Outcome measures	Number of studies	Blinded outcome assessment
Primary outcome measures:		
Pain*	15	7
– overall	12	6
– at rest	2	–
– under strain	1	–
– on activity	5	3
Global improvement	11	8
Overall severity of elbow complaints	1	–
Specific elbow function questionnaire*	2	1
– pain free function questionnaire	1	–
– function (VAS)	2	1
Secondary outcome measures:		
Grip strength*	19	15
– pain free grip strength	9	8
– maximum grip strength	3	2
– not specified	10	7
Patient satisfaction	3	–
Occupational status/sick leave	3	–
Weight test	3	2
Thermalgraphic gradient	1	1
Recreational status	1	–
Hobby limitations	1	–
Number of tender points	2	1
Pressure pain threshold	1	1
Middle finger test	2	1
Mill's test (number of positive tests)	1	–
Upper limb tension test IIb	2	1

* In some studies more than one method was used. Subcategories therefore add up to more than the total category.

 ISSN 0785-3890

Table 3. Methodological assessment of RCTs of physiotherapy for lateral epicondylitis*.

First author (reference)	Validity score†	Bias considered likely	Insufficient information is given for validity assessment	Incomplete information for description and data extraction
Vasseljen (45)	11	V11		D4, 7
Haker (31)	10	V3	V1	D2; S2
Haker (33)	10	V11	V1	S1, 2
Haker (34)	10	V11	V1	S2
Pienimäki (42)	10	V7, 8		D4, 5, 7
Lundeberg (40)	9		V1, 3, 5	D2, 4, 7
Papadopoulos (41)	9	V3, 5	V4	D1, 2, 4, 6, 7; S2
Haker (32)	8	V9, 11	V1, 3	D2; S2
Johannsen (37)	8	V11	V1, 9, 12	D3, 4, 6, 7; S1, 2
Vasseljen (44)	8	V7, 8, 11, 12		D4, 7
Binder (25)	7	V5	V1, 3, 4, 12	D1, 2, 4; S2
Krasheninnikoff (38)	7	V5, 11	V1, 4, 9	D4, 7; S1
Stratford (43)	7	V5, 7, 8	V3, 4	D2, 4, 6, 7; S1
Vicenzino (47)	7	V7	V1, 3, 4, 5	D2, 6, 7; S2
Verhaar (46,49)	6	V3, 5, 7, 8, 12	V4	D2, 6
Drechsler (28)	5	V7, 8	V1, 2, 3, 4, 12	D1, 2, 4, 7
Lundeberg (39)	5	V5	V1, 3, 4, 6, 9, 11	D2, 4, 7; S1
Chard (26)	4	V5	V1, 2, 3, 4, 6, 9, 11	D1, 2, 4, 7; S1, 2
Devereaux (27)	4	V5	V1, 3, 4, 6, 7, 9, 11	D1, 2, 7; S1, 2
Holdsworth (36)	4	V5, 7, 9, 11	V1, 3, 4, 8	D1, 2, 4, 6
Halle (35)	3	V5, 7, 8, 12	V2, 3, 6, 9, 11	D2, 3, 4, 6, 7; S2
Dwars (29)	1	V4, 5, 6, 7, 8, 9, 11	V1, 2, 3, 12	D1, 2, 3, 6, 7
Gudmundsen (30)	1	V8, 11	V1, 2, 3, 4, 5, 6, 7, 9, 10	D1, 2, 3, 4, 6, 7; S1, 2

* Ranked in order of internal validity score; equally ranked trials are ordered alphabetically. See Table 1 for definition of criteria numbers (V = validity; D = descriptive; S = statistical items); † validity score is calculated as the sum of all items with bias unlikely. Each item is given equal weight (range 0–12). Incomplete information for the validity assessment has a weight of 0.

likely, or for which insufficient information was given. Studies are ranked according to their internal validity score. Inter-rater agreement on the internal validity items was good (agreement 83%, kappa statistic 0.64) (21, 50). After consensus, 5 of the 27 disagreements could not be resolved, and a third reviewer (DW) made the final decision. Differences in interpretation of intention-to-treat analysis (V11) and dropouts (V9) caused most disagreements. Fourteen out of 23 studies had acceptable validity score (≥7 points).

Sample size and power

The sample sizes of the studies ranged between 5 and 60 patients per intervention group. None of the included studies were designed with sufficient power to detect an SMD of 0.5 (Table 4). Only three studies were designed with nearly sufficient power (0.78, 0.71, 0.72) to detect medium SMDs (0.5), but these studies had low validity scores (29, 30, 46).

Effectiveness of physiotherapy

Table 4 shows the effectiveness of each physiotherapy intervention on pain and global improvement. Except for ultrasound, statistical pooling was not possible because of a lack of similarity of study population, intervention, outcome measures, timing of outcome assessment, or insufficient reported data.

Laser. Eight studies (six with acceptable validity) compared the effects of laser with placebo (30–32, 34, 38, 39, 41, 45). Short-term follow-up (≤6 weeks) showed no statistically significant effects on pain. Contradictory results were reported for intermediate (6 weeks to 6 months) and long-term follow-up (≥6 months) assessments, and for comparisons with other physiotherapeutical modalities. Based on the best evidence synthesis there is insufficient evidence to demonstrate either benefit or lack of effect of laser for lateral epicondylitis.

Ultrasound. Three studies with acceptable validity, but low power, compared the effectiveness of ultrasound (US) with placebo (25, 33, 40). Two studies (25, 40) reported beneficial effects for ultrasound at 4, 8 and 13 weeks. Pooling of the intermediate term outcomes (25, 40) resulted in a large effect size for pain in favour of ultrasound (SMD −0.98 (−1.64, −0.33)). Based on the best evidence synthesis, there is weak evidence for the effectiveness of US in comparison with placebo. Seven studies compared US with other physiotherapy modalities or with other conservative treatments. Two studies of acceptable validity compared US with laser (44) or exer-

 ISSN 0785-3890

cises (42), and showed contradictory results. US with friction massage was superior to laser (SMD (95% CI) pain: −0.84 (−1.58, −0.09)) (49), but inferior to exercises (SMD (95% CI) pain: 0.95 (0.26, 1.64)) (42). Consequently, there appears to be insufficient evidence in favour of US when compared to other active interventions.

Electrotherapy. Only one (37) of four studies (26, 27, 35) evaluating the effectiveness of electrotherapy (electromagnetic field therapy, transcutaneous electral nerve stimulation) was of adequate methodological quality, but this study provided insufficient data on pain or general improvement to calculate an effect size. Due to a lack of information on outcomes, heterogeneity of interventions, low power, and poor quality of most studies, no conclusions can be drawn regarding the effectiveness of electrotherapy for lateral epicondylitis (insufficient evidence).

Exercises and mobilisation techniques. One study with acceptable validity demonstrated a large beneficial effect of exercises compared to ultrasound plus friction massage (SMD (95% CI): −0.95 (−1.64, −0.26)) (42). Four other studies were either of poor validity (28, 29, 46) or provided insufficient data on relevant outcome measures (47). The best evidence synthesis resulted in insufficient evidence for exercises and mobilisation techniques, because of low power, poor validity and large heterogeneity regarding interventions and outcomes.

Discussion

In this review, we assessed the methodological quality, results, and statistical power of 23 studies. The internal validity was acceptable in 14 studies. Based on our best evidence synthesis, weighting the studies with respect to their internal validity, power, statistically significant and clinically relevant results there was insufficient evidence to draw definite conclusions on the effects of laser, electrotherapy, exercises and mobilisation techniques for lateral epicondylitis. Weak evidence was found for the effectiveness of ultrasound compared to placebo.

Our review highlights a wide variation of interventions, study populations, and outcome measures used in lateral epicondylitis RCTs. These factors limit the degree to which the results of different studies can be compared and pooled. In addition, the heterogeneity of the timing and assessment of outcome measures, the inadequate reporting of results, contradictory results, low number of studies of each intervention, and the small sample sizes preclude making firm conclusions about the effectiveness of physiotherapy interventions. In addition, none of the studies were designed with sufficient power to detect a medium effect size (SMD of 0.5) and only seven studies had sufficient power to detect a large effect size (SMD of 0.8).

The question arises whether the contradictory treatment effects among studies are determined by clinical differences, methodological quality, or insufficient power. Schultz et al. (51) and Moher et al. (24) have demonstrated that inadequate trial methods, especially inadequate concealment of treatment allocation and lack of blinding, can result in biased estimates of the treatment effect. In addition, the variation of treatment effects will be larger among studies with small sample sizes, as demonstrated by our review.

Despite the clinical heterogeneity and overall inconclusive findings of this review, there are a few valid studies that suggest potential effectiveness of treatment. Firstly, ultrasound may be better than placebo in alleviating pain (25, 40). Secondly, ultrasound with friction massage may be more effective than laser treatment (44). Thirdly, stretching, muscle conditioning and occupational exercises may be better than ultrasound with friction massage (42). However, more research is definitely needed to strengthen or refute the evidence for the effectiveness of these interventions.

Most studies restricted their outcome assessment to pain, global improvement and grip strength, but a wide variety of outcome measures were used. The number of different outcome measures complicates the comparison of treatment effects, and the interpretation of effectiveness, because of differences in validity, reliability and responsiveness. In addition, most studies did not take into account the influence of potential prognostic variables such as concomitant neck disorders and the duration of elbow complaints. In only one study, a subgroup analysis for concomitant neck disorders was performed (46, 49).

Methodological remarks

Quality assessment can be performed using different checklists or scales. In our opinion the Amsterdam-Maastricht consensus list, which we used to assess methodological quality, represents a high standard for internal validity. Due to the unique challenges of adequately blinding patients and care provider in a trial of a physiotherapy intervention (as opposed to drugs trials) it is difficult to score highly on a tool such as the Amsterdam-Maastricht consensus list. To provide a comparison with the frequently used checklist of Jadad (18), our reviewers also scored the latter. Correlation between our internal validity score and the Jadad score was satisfactory (Spearman rank correlation coefficient 0.73).

Table 4. Summary of validity scores, sample size, power calculations and effect sizes * (95% confidence interval) for pain and global improvement for short, intermediate and long term outcome assessment for all included RCTs.

Intervention *Index versus reference group* First author (reference)	Validity score†	Sample size‡	**Short-term outcome assessment** (≤6 weeks) Week	Pain SMD	(95% CI)	Global improvement RR	(95% CI)	**Intermediate- and long-term outcome assessment** (>6 weeks) Week	Pain SMD	(95% CI)	Global improvement RR	(95% CI)	**Power§ to detect effect size of 0.5**	**Power to detect effect size of 0.8**
ULTRASOUND (US)														
US versus placebo														
Haker (33)	10	21	5	–		1.13	(0.68, 1.89)	13	–		–		0.35	0.72
								52	–		–			
Lundeberg (40)	9	33	5	–		0.79	(0.49, 1.27)	13	−1.33	(−1.87, −0.80)	0.91	(0.65, 1.29)	0.52	0.89
Binder (25)	7	38	4	−0.61	(−1.07, −0.15)	–		8	−0.66	(−1.13, −0.20)	0.52	(0.33, 0.82)	0.58	0.93
US versus no treatment														
Lundeberg (40)	9	33	5	–		0.63	(0.41, 0.96)	13	−1.70	(−2.26, −1.13)	0.44	(0.26, 0.74)	0.52	0.89
US versus US (different regimes)														
Holdsworth (36)	4	10	6	−0.77	(−1.79, 0.25)	−0.38	(−1.38, 0.61)¶	52	–		NA		0.19	0.40
Halle (35)	3	12	1	NA		–		–	–		–		0.22	0.47
US versus electrotherapy														
Halle (35)	3	12	1	NA		–		–	–		–		0.22	0.47
US + friction massage versus US														
Stratford (43) (placebo oinment)	7	9	5	−0.23	(−1.11, 0.66)	0.72	(0.43, 1.18)	–	–		–		0.17	0.36
Stratford (43) (hydrocortisone)	7	10	5	0.39	(−0.47, 1.26)	1.14	(0.69, 1.90)	–	–		–		0.19	0.40
US + friction massage versus exercises														
Pienimäki (42)	10	19	–	–		–		8	0.95	(0.26, 1.64)	–		0.32	0.67
US + friction massage versus laser														
Vasseljen (44)	8	15	3	−0.92	(−1.67, −0.17)	0.92	(0.62, 1.36)	7	−0.84	(−1.58, −0.09)	0.63	(0.26, 1.47)	0.26	0.56
US + friction massage + exercises versus mobilisation head radius														
Drechsler (28)	5	8	5	–		–		13	–		–		0.15	0.32
US versus injection														
Halle (35)	3	12	1	NA		–		–	–		–		0.22	0.47
LASER														
Laser versus placebo														
Vasseljen (45)	11	15	3	−0.25	(−0.96, 0.47)	0.81	(0.61, 1.06)	7	−0.46	(−1.19, 0.27)	0.67	(0.39, 1.14)	0.26	0.56
Haker (31)	10	23	5	–		1.45	(0.96, 2.20)	13	–		0.95	(0.51, 1.75)	0.38	0.76
								52	–		0.61	(0.22, 1.70)		
Haker (34)	10	29	4	–		0.87	(0.65, 1.16)	13	–		0.93	(0.56, 1.53)	0.46	0.85
								26	–		0.46	(0.23, 0.93)		
								52	–		–			

Papadopoulos (41)	9	14	2	NA		–		–	–		–		0.25	0.53
Haker (32)	8	26	5	–		–		13	–		–		0.42	0.81
								52	–		–			
Krasheninnikoff (38)	7	8	4	NA		1.07	(0.82, 1.39)	14	NA		1.00	(0.63, 1.59)	0.15	0.32
Lundeberg (39)	5	19	2	NA		–		13	−2	(−2.77, −1.22)	–		0.32	0.67
								26	–		NA			
Gudmundsen (30)	1	52	4	–		0.72	(0.60, 0.87)	–	–		–		0.71	0.98
Laser versus US + friction massage														
Vasseljen (44)	8	15	3	0.92	(0.17, 1.67)	1.09	(0.73, 1.62)	7	0.84	(0.09, 1.58)	1.60	(0.68, 3.77)	0.26	0.56
Laser versus laser														
Lundeberg (39)	5	19	2	NA		–		13	−1.00	(−1.67, −0.33)	–		0.32	0.67
								26	–		NA			
ELECTROTHERAPY														
Electrotherapy versus placebo														
Johannsen (37)	8	7	4	NA		–		–	–		–		0.14	0.28
Chard (26)	4	17	2, 4, 6	–		NA		8	–		0.38	(0.12, 1.18)	0.29	0.62
Devereaux (27)	4	15	6	–		–		8	–		–		0.26	0.56
Electrotherapy versus US														
Halle (35)	3	12	1	NA		–		–	–		–		0.22	0.47
Electrotherapy versus injection														
Halle (35)	3	12	1	NA		–		–	–		–		0.22	0.47
EXERCISES AND MOBILISATION TECHNIQUES														
Cervical manipulation versus placebo														
Vicenzino (47)	7	5	1 day	NA		–		–	–		–		0.11	0.20
Cervical manipulation versus no treatment														
Vicenzino (47)	7	5	1 day	NA		–		–	–		–		0.11	0.20
Stretching + strengthening exercises versus US + friction massage														
Pienimäki (42)	10	19	–	–		–		8	−0.95	(−1.64, −0.26)	–		0.32	0.67
Mobilisation head radius versus US + friction massage + exercises														
Drechsler (28)	5	8	5	–		–		13	–		–		0.15	0.32
Friction massage + Mills' manipulation versus injection														
Verhaar (46,49)	6	53	6	1.63	(1.28, 2.08)‖	2.22	(1.45, 3.39)	52	0.83	(0.66, 1.04)‖	0.82	(0.54, 1.26)	0.72	0.98
Friction massage + exercises versus elbow support														
Dwars (29)	1	60	6	−0.17	(−0.60, 0.27)	0.97	(0.61, 1.56)	–	–		–		0.78	0.99

* Values are the effect sizes (95% confidence intervals) for each outcome measure; standardised mean differences (SMD) for continuous outcome measures and relative risks (RR) for dichotomous outcome measures. SMD < 0 in favour of index group; RR < 1 in favour of index group; † validity score: range 0–12, higher values indicating less potential bias; ‡ smallest group; § Based on a t-test for differences between the means of 2 independent samples of equal size and equal variances; power ≥0.80 is generally considered sufficient ($\alpha = 0.05$); NA = not able to calculate effect sizes due to insufficient data presentation; ‖ analysed as continuous data (SMD). | analysed as dichotomous data (RR).

The conclusions of our review were dependent on few, important decision rules. These were our quality assessment (52), and the cut-off points used in our best-evidence synthesis (e.g., acceptable versus low internal validity (12), clinically relevant results (23), number of studies needed for the qualitative analysis, consistency of finding of individual studies, power (53), and conclusions on the strength of evidence for physiotherapy). These decision rules were arbitrary. However, we also performed sensitivity analyses, which included studies with low internal validity in both the quantitative and qualitative analyses (data not shown). These did not substantially alter our conclusions.

Implications for research

Disappointingly, nine years and an additional 18 RCTs later, our review can only confirm the conclusions by Labelle et al. (6). In our opinion, more research into the effectiveness of physiotherapy for lateral epicondylitis is definitely still needed. Such trials should have larger sample sizes, include a 'no treatment group' or 'placebo treatment', use clinically appropriate valid, reliable and responsive outcome measures and have a short- ($\leq$6 weeks), intermediate- (6 weeks to 6 months) and long-term follow-up ($\geq$6 months). In addition, results in subgroups based on important prognostic factors (e.g., duration of elbow complaints and concomitant neck complaints) should be presented separately. More pragmatic trials are needed to directly compare common interventions for lateral epicondylitis, as they are carried out in everyday care, including combinations of physiotherapy intervention (e.g., ergonomic advice, elbow mobilisation, deep friction massage, ultrasound, exercises programme and an additional home exercise programme).

The reliability and responsiveness of the outcome measures used in most trials are unknown. To improve our ability to interpret and compare the results of different studies, further work is needed to standardise measures and address measurement issues. Incorporation of a standard set of outcome measures in all future RCTs on epicondylitis may greatly enhance these efforts (54–57).

In conclusion, there is little evidence to support or refute the use of physiotherapy for lateral epicondylitis. The lack of evidence, however, is mainly the result of low statistical power, inadequate validity and insufficient reported data of the majority of trials. Additional well-designed pragmatic trials are needed to provide evidence of value of common physiotherapy interventions for lateral epicondylitis.

The authors would like to thank Hans Lund, Tatjana Rogic and Marco Romoli for their help with the selection of RCTs, quality assessment and data-extraction of the Norwegian, Serbo Croatian and Italian articles, Samyra Keus for blinding the articles, and the Cochrane Musculoskeletal Group Editorial Team for their helpful comments in reviewing the protocol of the review.

Contributors: NS coordinated the review under WJJA supervision. NS planned and initiated the review, searched and selected the literature, assessed the studies, retrieved data, analysed and interpreted the results. WJJA selected the literature. HA assessed the studies and retrieved data. DAWMW assessed the studies. All authors contributed to the composition of the paper.

Funding: Dutch Organisation for Scientific Research (NWO) and the Dutch Health Insurance Council.

Conflict of interest: none.

This review will be published in a more extensive version and regularly updated in the Cochrane Library (14).

References

1. **Allander E.** Prevalence, incidence and remission rates of some common rheumatic diseases and syndromes. *Scand J Rheumatol* 1974; 3: 145–53.
2. **Chard MD, Hazleman BL.** Tennis elbow — a reappraisal. *Br J Rheumatol* 1989; 28: 187–90.
3. **Chop WM.** Tennis elbow. *Postgrad Med* 1989; 86: 307–8.
4. **Murtagh JE.** Tennis elbow. *Aust Fam Physician* 1988; 17: 90–1, 94–5.
5. **Beckerman H, Bouter LM, Heijden GJ van der, Bie RA de, Koes BW.** Efficacy of physiotherapy for musculoskeletal disorders: what can we learn from research? *Br J Gen Pract* 1993; 43: 73–7.
6. **Labelle H, Guibert R, Joncas J, Newman N, Fallaha M, Rivard C-H.** Lack of scientific evidence for the treatment of lateral epicondylitis of the elbow: an attempted meta-analysis. *J Bone Joint Surg* 1992; 74B: 646–51.
7. **Ernst E.** Mini review: Use a new treatment while it still works: ultrasound for epicondylitis. *Eur J Phys Med* 1994; 4: 50–1.
8. **Windt DAWM van der, Heijden GJ van der, Berg SGM van den, Riet G ter, Winter de AF, Bouter LM.** Ultrasound therapy for musculoskeletal disorders: a systematic review. *Pain* 1999; 81: 257–71.
9. **Hunt DL, McKibbon KA.** Locating and appraising systematic reviews. *Ann Intern Med* 1997; 126: 532–8.
10. **Gregoire G, Derderian F, Lorier J le.** Selecting the language of the publications included in a meta-analysis: is there a tower of babel bias? *J Clin Epidemiol* 1995; 48: 158–63.
11. **Egger M, Smith GD.** Meta-analysis: potentials and promise. *BMJ* 1997; 315: 1371–4.
12. **Detsky AS, Naylor CD, O'Rourke K, McGeer AJ, L'Abbe KA.** Incorporating variations in the quality of individual randomized trials into meta-analysis. *J Clin Epidemiol* 1992; 45: 255–65.
13. **Moher D, Fortin P, Jadad AR, Juni P, Klassen T, Le Lorier J, et al.** Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996; 347: 363–6.
14. **Mulrow CD, Oxman AD, eds.** Cochrane Collaboration Handbook (updated September 1997). In: *The Cochrane*

Library (database on disk and CDROM). The Cochrane Collaboration. Oxford: Update Software 1999; Issue 1.
15. **Tulder MW van, Assendelft WJJ, Koes BW, Bouter LM.** Method guidelines for systematic reviews in: The Cochrane Collaboration Back Review Group for Spinal Disorders. *Spine* 1997; 22: 2323–30.
16. **Schulz KF, Chalmers I, Grimes DA, Altman DG.** Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynaecology journals. *JAMA* 1994; 272: 125–8.
17. **Heijden van der GJ, Windt van der DA, Winter de AF.** Physiotherapy for patients with soft tissue shoulder disorders: a systematic review of randomised clinical trials. *BMJ* 1997; 315: 25–30.
18. **Jadad AR, Moore A, Carroll D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al.** Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17: 1–12.
19. **Verhagen AP, Vet HCW de, Bie RA de, Kessels AGH, Boers M, Bouter LM, et al.** The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998; 5: 1235–41.
20. **Cohen J.** A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.
21. **Brennan P, Silman A.** Statistical methods for assessing observer variability in clinical measures. *BMJ* 1992; 304: 1491–4.
22. **Anonymous.** *Wissenschaftliche tabellen, Teilband statistik*. 8th edition. Basel: Geigy; 1980.
23. **Cohen J.** *Statistical power analysis for the behavioral sciences*. Hillsdale (NJ): Lawrence Erlbaum Associates; 1988.
24. **Moher D, Jones D, Cook DJ, Jadad AR, Moher M, Tugwell P, et al.** Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998; 352: 609–13.
25. **Binder A, Hodge G, Greenwood AM, Hazleman BL, Page Thomas DP.** Is therapeutic ultrasound effective in treating soft tissue lesions? *BMJ* 1985; 290: 512–4.
26. **Chard MD, Hazleman BL.** Pulsed electromagnetic field treatment of chronic lateral humeral epicondylitis. *Clin Exp Rheumatol* 1988; 6: 330–6.
27. **Devereaux MD, Hazleman BL, Thomas PP.** Chronic lateral humeral epicondylitis — a double-blind controlled assessment of pulsed electromagnetic field therapy. *Clin Exper Rheumatol* 1985; 3: 333–6.
28. **Drechsler WI, Knarr JF, Snyder-Mackler L.** A comparison of two treatment regimens for lateral epicondylitis: a randomized trial of clinical interventions. *J Sport Rehabil* 1997; 6: 226–34.
29. **Dwars BJ, Feiter P de, Patka P, Haarman HJThM.** Functional treatment of tennis. *Sports Medicine and Health* 1990: 237–41.
30. **Gudmundsen J, Vikne J.** Laserbehandling av epicondylitis humeri og rotatorcuffsyndrom. *Nor Tidskr Idrettsmed* 1987; 2: 6–15.
31. **Haker E, Lundeberg T.** Laser treatment applied to acupuncture points in lateral humeral epicondylalgia. A double-blind study. *Pain* 1990; 43: 243–7.
32. **Haker E, Lundeberg T.** Is low-energy laser treatment effective in lateral epicondylalgia? *J Pain Symptom Manage* 1991; 6: 241–6.
33. **Haker E, Lundeberg T.** Pulsed ultrasound treatment in lateral epicondylalgia. *Scand J Rehabil Med* 1991; 23: 115–8.
34. **Haker EH, Lundeberg TC.** Lateral epicondylalgia: report of noneffective midlaser treatment. *Arch Phys Med Rehabil* 1991; 72: 984–8.
35. **Halle JS, Franklin RJ, Karalfa L.** Comparison of four treatment approaches for lateral epicondylitis of the elbow. *J Orthop Sport Phys Med* 1986; 8: 62–9.
36. **Holdsworth LK, Anderson DM.** Effectiveness of ultrasound used with a hydrocortisone coupling medium or epicondylitis clasp to treat lateral epicondylitis: pilot study. *Physiotherapy* 1993; 79: 19–25.
37. **Johannsen F, Gam A, Hauschild B, Mathiesen B, Jensen L.** Rebox: an adjunct in physical medicine? *Arch Phys Med Rehabil* 1993; 74: 438–40.
38. **Krasheninnikoff M, Ellitsgaard N, Rogvi-Hansen B, Zeuthen A, Harder K, Larsen R, et al.** No effect of low power laser in lateral epicondylitis. *Scand J Rheumatol* 1994; 23: 260–3.
39. **Lundeberg T, Haker E, Thomas M.** Effect of laser versus placebo in tennis elbow. *Scand J Rehabil Med* 1987; 19: 135–8.
40. **Lundeberg T, Abrahamsson P, Haker E.** A comparative study of continuous ultrasound, placebo ultrasound and rest in epicondylalgia. *Scand J Rehabil Med* 1988; 20: 99–101.
41. **Papadopoulos ES, Smith RW, Cawley MID.** Low-level laser therapy does not aid the management of tennis elbow. *Clinical Rehabil* 1996; 10: 9–11.
42. **Pienimäki T, Tarvainen TK, Siira PT, Vanharanta H.** Progressive strengthening and stretching exercises and ultrasound for chronic lateral epicondylitis. *Physiotherapy* 1996; 82: 522–30.
43. **Stratford PW, Levy DR, Gauldie S, Miseferi D, Levy K.** The evaluation of phonophoresis and friction massage as treatments for extensor carpi radialis tendinitis: a randomized controlled trial. *Physiotherapy Canada* 1989; 41: 93–9.
44. **Vasseljen O.** Low-level laser versus traditional physiotherapy in the treatment of tennis elbow. *Physiotherapy* 1992; 78: 329–34.
45. **Vasseljen O, Hoeg N, Kjeldstad B, Johnsson A, Larsen S.** Low level laser versus placebo in the treatment of tennis elbow. *Scand J Rehabil Med* 1992; 24: 37–42.
46. **Verhaar JAN, Walenkamp GH, Mameren H van, Kester AD, Linden AJ van der.** Local corticosteroid injection versus Cyriax-type physiotherapy for tennis elbow. *J Bone Joint Surg* 1996; 78B: 128–32.
47. **Vicenzino B, Collins D, Wright A.** The initial effects of a cervical spine manipulative physiotherapy treatment on the pain and dysfunction of lateral epicondylalgia. *Pain* 1996; 68: 69–74.
48. **Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, et al.** Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; 354: 1896–900.
49. **Verhaar JAN.** *Tennis elbow* (thesis). Maastricht: University Press, 1992.
50. **Altman DG.** *Practical Statistics for Medical Research*. Cornwall: Chapman and Hall; 1991: 404.
51. **Schultz KF, Chalmers I, Hayes RJ, Altman DG.** Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995; 273: 408–12.
52. **Juni P, Witschi A, Bloch R, Egger M.** The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999; 282: 1054–60.
53. **Moher D, Dulberg CS, Wells GA.** Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994; 272: 122–4.
54. **Wolfe F, Lassere M, Heijde D van der, Stucki G, Suarez-Almazor M, Pincus T, et al.** Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology. *J Rheumatol* 1999; 26: 484–9.
55. **Deyo RA, Battie M, Beurskens AJHM, Bombardier C, Croft P, Koes B, et al.** Outcome measures for low back pain

 ISSN 0785-3890

research. A proposal for standardized use. Spine 1998; 23: 2003–13.

56. **Green S, Buchbinder R, Glazier R, Forbes A.** Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. *BMJ* 1998; 316: 354–60.

57. **Boers M, Tugwell P, Felson DT, Riel van PLCM, Kirwan JR, Edmonds JP, et al.** World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials. *J Rheumatol* 1994; 21 Suppl 41: 86–9.

 ISSN 0785-3890